

INT-2024

~~CEPAL/CELADE(2024)~~



BANCO INTERAMERICANO DE DESARROLLO
CENTRO LATINOAMERICANO DE DEMOGRAFIA



**Jerarquización de las provincias de Perú
según grado de pobreza: aspectos metodológicos***

* versión preliminar elaborada por el consultor de largo plazo del Convenio BID/CELADE. Sr. Jorge Bravo.

Junio, 1995

Programa Regional para la Incorporación del Análisis Demográfico en los Proyectos de Inversión y para la Capacitación en el Uso de la Información Demográfica en la Planificación, Preparación y Evaluación de Proyectos (ATN/TF-4098-RG).

1. Antecedentes y Objetivo

Este documento se elabora como un aporte del Convenio BID/CELADE a los programas de inversión social que el Banco Interamericano de Desarrollo (BID) apoya en Perú. El documento persigue, más particularmente, presentar una propuesta metodológica para integrar información socio-demográfica de censos y encuestas de Perú, con el propósito de obtener indicadores de pobreza a nivel de provincias que sirvan a una más eficiente asignación de recursos en las políticas y los programas sociales.

Uno de los principios que guían el diseño y la ejecución de dichas políticas y programas, sobre todo aquellas que tienen un alcance no-universal o focalizado, es la *priorización* de la asignación de recursos. Esta priorización se establece, de modo importante, en función del grado de necesidad o de carencias que afectan a los diferentes grupos que componen la población; es decir, en función de su grado de pobreza. Una de las dimensiones importantes de dicha priorización es la geográfica o político-administrativa, en atención a la funcionalidad que ella representa para la ejecución de programas. Quedan comprendidos en tal caso, por ejemplo, los programas o fondos de inversión social que funcionan en muchos países de la región, y algunos proyectos de emergencia, como el de alivio y reducción de la pobreza en Perú, iniciado el año 1994. En general, la jerarquización de las unidades político-administrativas es relevante a todo programa orientado a aliviar las condiciones de pobreza más críticas.

Sin embargo, al intentar operacionalizar dicho principio y obtener indicadores sintéticos apropiados, se encuentran algunas dificultades prácticas. Por un lado, los datos que tienen un adecuado nivel de desagregación son generalmente insuficientes. Estos datos son, típicamente, indicadores de necesidades básicas insatisfechas provenientes del censo. Si bien se encuentra allí una cantidad de información relevante a niveles geográficos

suficientemente desagregados, no resulta fácil obtener un indicador sintético de aceptación generalizada por parte de los analistas o los tomadores de decisiones. En particular, el censo de 1993, así como la gran mayoría de los censos en la región de América Latina y el Caribe, no recogen información de ingresos, lo que impide una estimación directa de la proporción de hogares debajo de la línea de pobreza, que es un indicador síntesis de relativa amplia aceptación y utilización. Por otro lado, las encuestas nacionales de hogares y de características socioeconómicas contienen información suficientemente detallada de ingresos para hacer las estimaciones requeridas, pero su representatividad alcanza, sólo a grandes regiones, áreas urbanas o aún metropolitanas.

En una primera mirada, pues, y en la ausencia de otras estadísticas completas y confiables al nivel requerido, el problema pareciera ser insalvable. Sin embargo, tal como sugiere este documento, es posible obtener estimaciones *indirectas* de la distribución del ingreso de los hogares mediante la combinación de ambas fuentes de información (censo y encuestas) y del uso de métodos estadísticos apropiados. Un antecedente de que dicha combinación es no tan sólo factible, sino también útil desde el punto de vista de las políticas sociales, lo constituye el cruce de información de ingresos y de necesidades básicas insatisfechas, expuesto en la extensamente aplicada aproximación de Rubén Kaztman (1989) para la caracterización de la pobreza.

Partiendo de la posibilidad de combinar los *tipos* de información antedichos, estas notas exponen brevemente una posible forma de integrar las *fuentes* de información (léase el censo de 1993 y la "Encuesta Nacional de Propósitos Múltiples", ENAPROM)- para obtener ciertos indicadores de nivel de vida a nivel de departamentos y provincias. Es importante anotar que la ENAPROM cubría, hasta 1994, sólo el área metropolitana de Lima y Callao y las 24 ciudades capitales del resto de los departamentos del país. A partir de 1995, la encuesta de hogares de Perú ha extendido su cobertura a la totalidad

del país, con representatividad a nivel nacional, área urbana y rural. A partir de 1995 la encuesta se denomina Encuesta Nacional de Hogares (ENAHO).

La idea básica del método aquí propuesto es realizar una estimación indirecta del nivel de vida en cada provincia, expresado a través del porcentaje de pobreza, en dos etapas:

1. Establecer una relación estadística entre el nivel de ingreso de los hogares y un conjunto de características de las personas, de la vivienda y del tipo de localidad donde ellas habitan;
2. Usar el modelo desarrollado para estimar el porcentaje de pobreza a nivel departamental y provincial sobre la base de las características ya señaladas. Dicho indicador puede servir para el objetivo deseado de jerarquización de distritos.

2. Metodología

2.1 Factibilidad de combinar fuentes.

Esta sección examina brevemente la posibilidad práctica de combinar información sobre los hogares y sus miembros del Censo de 1993 y la ENAPROM. Tal como sugiere el cuadro 1, hay una cantidad apreciable de variables que fueron recogidas tanto en el censo como en la ENAPROM. Existen varias excepciones, como la información sobre los materiales de paredes, techo y piso, que fueron recopiladas en el censo pero no en la ENAPROM, así como algunas variables específicas que fueron captadas en la ENAPROM pero no en el censo (p.ej., combustible para cocinar, porqué no asiste a la escuela, preguntas acerca de actividades secundarias y otras). Con todo, sin embargo, el conjunto de variables comunes es bastante amplio para los propósitos presentes. Además, la gran mayoría de las variables en común tienen un formato similar y códigos compatibles. Lo anterior, sumado al hecho de que los cuestionarios son bastante coherentes entre sí y que ambas encuestas se llevan bajo la responsabilidad de una sola institución,

Cuadro 1. Información sobre hogares en Perú 1993, según fuente

Módulo/tema/pregunta	códigos		ENAPROM
	Censo	compatibles?	
Localización	✓	si	✓
Vivienda			
tipo (particular)	✓	si	✓
propiedad/uso	✓	si	✓
material paredes	✓		-
material techo	✓		-
material piso	✓		-
abastecimiento agua	✓	si	✓
baños	✓	si	✓
alumbrado eléctrico	✓	si	✓
combustible cocinar	-		✓
no. de habitaciones	✓	si	✓
no. de dormitorios	✓	si	✓
cocina	✓		-
producción de ingresos en el hogar	✓		-
equipamiento	✓	si	✓
Personas			
Educación			
no. de personas	✓	si	✓
parentesco	✓	si	✓
permanencia	-		✓
sexo	✓	si	✓
edad	✓	si	✓
estado civil	✓	si	✓
religión	✓		-
idioma materno	✓		-
alfabetismo	✓	si	✓
asiste	✓	no	✓
último grado	✓	no	✓
porqué no asiste	-		✓
Ocupación y Empleo			
profesión u oficio	✓		-
actividad semana pasada	✓	si	✓
ocupación de la s.p.	✓	si	✓
rama de actividad	✓	si	✓
categoría	✓	si	✓
no. de trabajadores			
tipo, lugar, horas	-		✓
actividad principal y secundaria	-		✓
Salud	-		✓
Migración	-		✓
Gastos del Hogar	-		✓
Transferencias realizadas	-		✓
Transacciones financieras	-		✓
Ingresos	-		✓

Fuente: INEI (1993a) Cédula Censal; INEI (1993b) Cuestionario ENAPROM-93

el INEI, hacen presumir que la información recogida debiera ser suficientemente comparable.

2.2. Modelo estadístico

Ya se señaló que un antecedente útil relacionado con lo aquí expuesto son los métodos que usan información de encuestas en cruces de necesidades básicas insatisfechas con líneas de pobreza. La propuesta presente persigue extender esa ligazón a la vinculación entre censo y encuesta, de modo que permita jerarquizar las provincias según indicadores del nivel de vida.

Para efectos del modelo aquí propuesto, se parte de la premisa general de que existe una relación entre el nivel de ingreso de un hogar (Y) y el número de miembros económicamente activos (n), sus características individuales (i), las características de la vivienda (v) y la zona de residencia (z) dentro del país. Se propone, como primer paso, la estimación de un modelo de regresión de la variable dependiente Y respecto de las variables independientes n, i, v, z , sobre la base de los datos de la ENAHO-1995, y ajustar el modelo a nivel provincial con base en las variables independientes del Censo-1993 (INEI, 1994). Un procedimiento análogo se llevó a cabo para efectos de estratificación en una encuesta de hogares en Paraguay (Sauma, 1993; anexo 3), con el objeto de imputar el nivel relativo de ingreso en 1982 a partir de la información censal de ese año y la relación estadística entre ingreso en 1990 y variables no-ingreso en ese mismo año.

La lista preliminar de variables independientes debe incluir a todas aquellas que sean comunes, ya sea exacta o aproximadamente, a la ENAHO y al Censo. Una posible especificación de las variables es la siguiente:

Declaración
Pablo

- n* - número de miembros Económicamente Activos (EA)

- i* - edad promedio de los EA en el hogar
 - nivel educativo promedio de los EA en el hogar
 - tipo de ocupación de los EA en el hogar
 - rama de actividad de los EA en el hogar

- v* - tipo de vivienda
 - abastecimiento de agua
 - tipo de servicios sanitarios
 - disponibilidad de electricidad
 - hacinamiento
 - disponibilidad de vehículos
 - equipamiento del hogar

- z* - zona de residencia (área metropolitana, ciudades grandes, ciudades medianas y pequeñas, áreas rurales)

Nótese que algunas de estas variables deben ser calculadas o codificadas previamente a la realización de regresiones propuestas en la sección siguiente. Por ejemplo, el tipo de ocupación y la rama de actividad deberán agruparse en unas pocas categorías cada una y definir un índice (p.ej., un promedio ponderado) de composición de ocupaciones y rama. La versión más simple de este procedimiento sería tomar, en relación con la ocupación, la proporción de los miembros EA en ocupaciones pertenecientes a los grupos principales 0 a 5 de la clasificación del Censo 1993, que incluye a los miembros de las fuerzas armadas, miembros del poder ejecutivo, profesionales, técnicos, jefes y empleados de oficina, y a trabajadores calificados. Este grupo está más estrechamente

asociado al sector "moderno" del mercado laboral y. es de presumir, podría determinar una relación distinta entre ingreso del hogar y las variables de vivienda, las individuales y las de zona de residencia ya mencionadas.

Consideraciones y definiciones análogas deberán realizarse respecto de todas las demás variables categóricas. Especial mención cabe, entre éstas, a la variable z , referida a la zona de residencia.

De definirse la variable dependiente como el ingreso *per capita* en lugar del ingreso total del hogar, parece razonable redefinir la variable n como el cociente del número de miembros económicamente activos entre el número total de personas en el hogar.

Como un paso previo a cualquier análisis estadístico, sería altamente conveniente una evaluación de la variable ingreso, especialmente en lo relativo a posibles problemas de sub-declaración. Un procedimiento comunmente utilizado en análisis de pobreza es tomar como referencia las cuentas de ingresos y gastos de las cuentas nacionales y realizar ajustes diferenciados según tipo de ingreso.

2.3. Ajuste y Estimación

Recuérdese que el objetivo final del trabajo es obtener una medida de la proporción (P) de los hogares de cada distrito que se ubican por debajo de la línea de pobreza. El primer paso propuesto es ajustar a los datos de la ENAHO-95 un modelo general a nivel de hogar del tipo:

$$Y_h = Y(n_h, i_h, v_h, \varepsilon_h) \quad (1)$$

donde h denota al hogar de referencia y ε_h al término de error del ingreso del hogar, ó bien uno del tipo:

$$\pi_h = \pi(n_h, i_h, v_h, \mu_h) \quad (2)$$

donde π_h denota la probabilidad de que el hogar h se sitúe bajo la línea de pobreza y μ_h al término de error de esta ecuación. Se deberá prestar atención a la especificación, ajuste y validación del modelo de regresión. Habrá que ensayar con diferentes formas funcionales en las variables independientes, así como detectar y minimizar posibles problemas de multicolinealidad, especialmente en el conjunto de variables v .

Una vez determinado el modelo de regresión final, existen dos posibilidades para estimar P : i) se usan los coeficientes estimados de la ecuación (1) y la información del censo correspondiente a las variables independientes de cada *hogar* y se hace una estimación de su nivel de ingreso o de su ingreso per cápita. Para cada provincia, se calcula el porcentaje de pobreza sobre la base de los valores estimados de Y_h ; ii) se usan los coeficientes estimados de la ecuación (2) y la información del censo correspondiente a las variables independientes de cada *provincia* y se calcula directamente el valor ajustado de P para la provincia. Cada una de estas variantes tiene ventajas y desventajas (véase el Apéndice); eventualmente, podrían implementarse ambas para evaluar la consistencia de los ordenamientos resultantes.

Finalmente, sería deseable realizar algún análisis de consistencia respecto de los resultados globales de la ENAHO y de sensibilidad respecto de la especificación del modelo estadístico, que permita evaluar la robustez de la jerarquización resultante.

3. Comentarios y Sugerencias

Como puede deducirse de las secciones anteriores, la combinación de información de la encuesta nacional de hogares y del Censo es factible y -mediante un trabajo estadístico cuidadoso- puede resultar de mucha utilidad para basar una jerarquización de

departamentos y provincias según nivel de vida. Que la ENAHO que recoja una cantidad apreciable de información del mismo carácter y en un formato similar al censo y que tenga representatividad a nivel nacional, urbano y rural, son factores que permiten tener optimismo respecto de la utilidad y confiabilidad de los resultados.

Un primer acercamiento a la jerarquización podría realizarse sobre la base de los indicadores censales recientemente publicados (INEI, 1994) a nivel departamental y provincial. Como un primer punto de comparación y posterior verificación de consistencia, parece conveniente obtener un primer conjunto de ordenamientos uni ó multivariados con la información censal solamente. Los componentes principales de la matriz $\{n i v z\}$ o mediante análisis de discriminantes podrían ser particularmente útiles a este efecto. La preparación de un archivo de las variables ya mencionadas a nivel de hogar será, en cualquier caso, requerido para la segunda etapa propuesta, por lo que ese análisis preliminar será complementario a las fases siguientes del trabajo.

Apéndice

Tal como se señala en la página 7, en el contexto presente existen dos formas de obtener una estimación de la proporción (P) de los hogares de cada provincia que se ubican por debajo de la línea de pobreza: **i)** realizar un ajuste (mediante la ecuación 1) del ingreso de cada hogar en el distrito y usar la distribución de hogares según nivel de ingreso para calcular P , ó **ii)** realizar un ajuste (mediante la ecuación 2) de la probabilidad π_h de que el hogar h se sitúe bajo la línea de pobreza y utilizar el valor ajustado de π para cada provincia como la estimación de P .

El procedimiento **i** tiene como ventaja que su interpretación conceptual es bastante directa y, adicionalmente, entrega una distribución de los hogares según ingreso, la que se puede utilizar para hacer diferentes cortes, además del de la línea de pobreza aquí privilegiado. Tiene como desventaja un requerimiento de un gran volumen de cálculos, ya que la estimación se hace sobre los más de 4 millones 760 mil hogares relevados en el censo de julio de 1993. Otra desventaja es que no existe garantía, *a priori*, de que algunos de los valores estimados se ubiquen en un rango fuera de los márgenes observados en la ENAPROM.

El procedimiento **ii**, por su parte, tiene como ventaja que requiere de una cantidad de cálculos muchísimo menor (sobre las 187 provincias del país), además que los programas estadísticos disponibles incorporan la restricción deseable de que los valores estimados de π y P se ubiquen dentro del rango (0,1). Tiene el pequeño inconveniente, más bien de carácter conceptual, de que involucra usar estimaciones hechas sobre la base de una variable dicotómica (que asume el valor 1 si el hogar está bajo la línea de pobreza, 0 si no lo está) para predecir el valor de una variable continua (la proporción de hogares o de población bajo la línea de pobreza). Ello requiere realizar alguna presunción respecto de la forma de la distribución probabilística de π . Este último inconveniente, sin embargo, puede superarse en buena medida experimentando con diferentes funciones de distribución para π , de modo de minimizar el riesgo de sesgo debido a la elección de una función de distribución muy específica o poco flexible.

Referencias

- INEI (1993a) *Cédula Censal*, Censos Nacionales 1993: IX de Población y IV de Vivienda. Lima, Perú.
- INEI (1993b) *Cuestionario ENAPROM-1993*. Encuesta Nacional de Propósitos Múltiples. Lima, Perú.
- INEI (1994) *Mapa de Necesidades Básicas Insatisfechas*. Lima, Perú.
- Kaztman, R. (1989) La heterogeneidad de la pobreza: el caso de Montevideo. *Revista de la CEPAL* No. 37, abril 1989, pp. 141-152.
- Sauma, P. (1993) *La distribución del ingreso en el Paraguay*. Facultad de Ciencias Económicas, Administrativas y Contables, Universidad Nacional de Asunción. Proyecto BID ATP/SF-2156-PR.